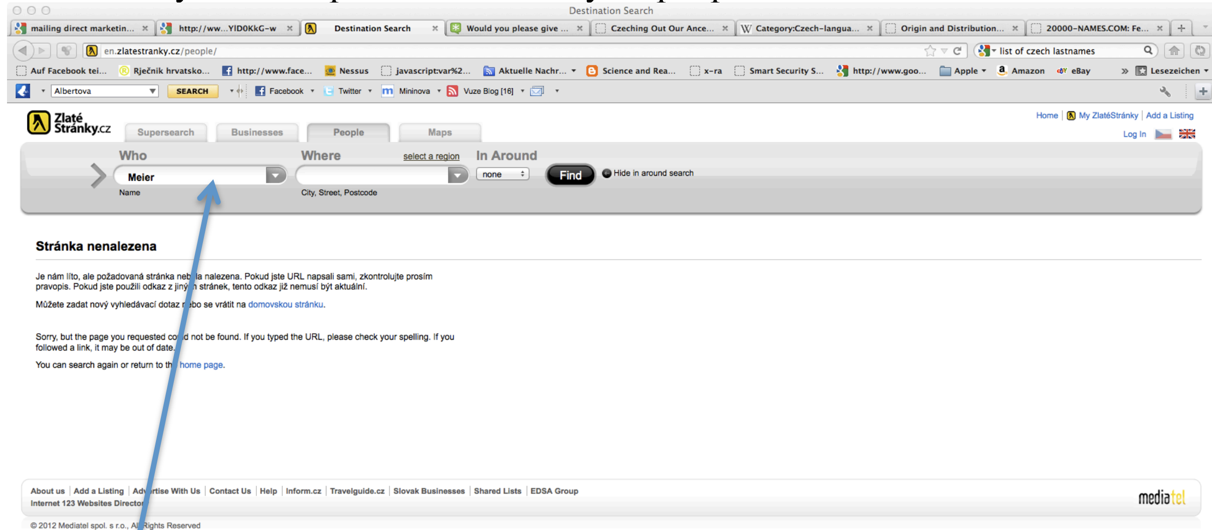
Addresscollector from Zlate Strank in Czech Republic

The script should crawl all possible combinations of names on the webpage of Zlatestranky: http://en.zlatestranky.cz/people/



Here is the lastname and search parameter entere das well as the city...

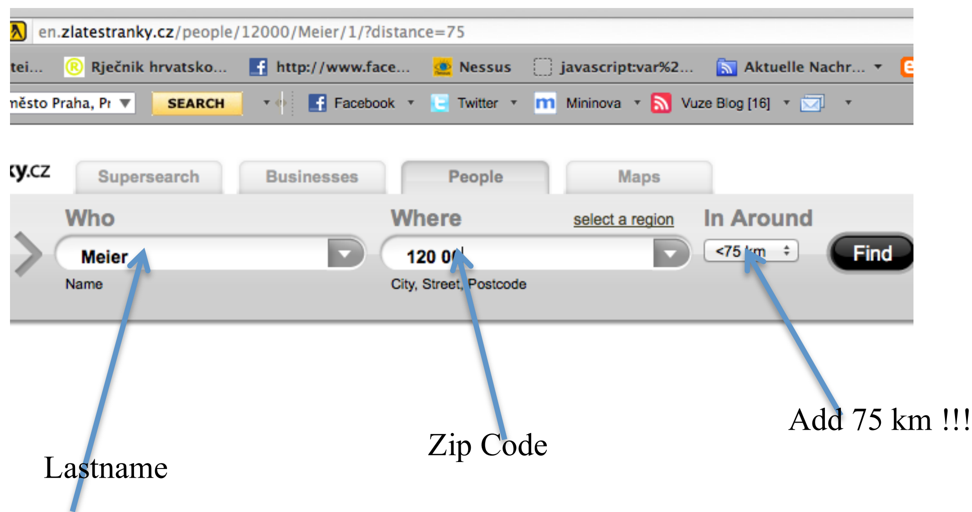
You can find a list of all Czech Surnames here: http://www.iarelative.com/czech/a_b.htm
http://www.iarelative.com/czech/alpha.htm
http://en.wikipedia.org/wiki/Category:Czech-language_surnames

---

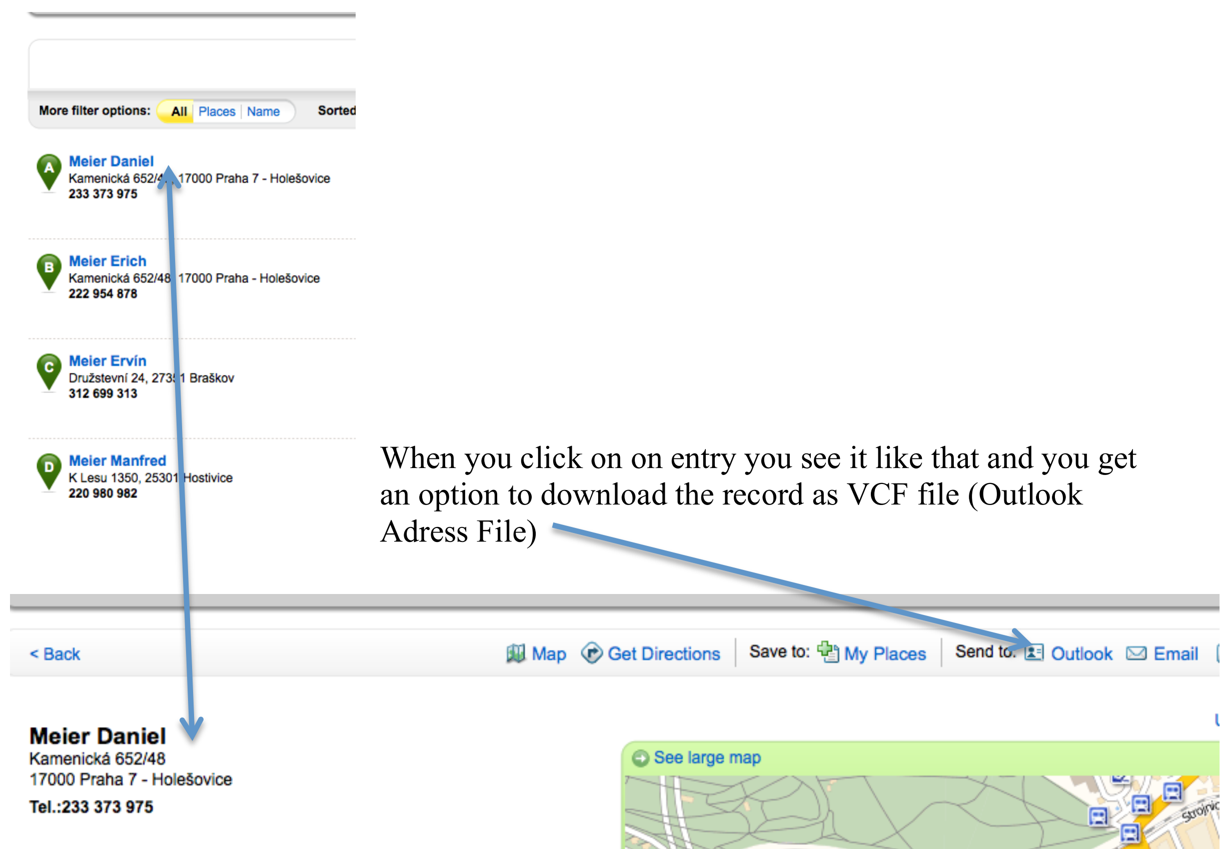
You should write the script like that, that you check all lastnames in every ZIP code, where you check in 10000, 11000, 12000, 13000, ...till , 99000 where you add the extension that you search in an area of 75 kilometer:

http://en.zlatestranky.cz/people/12000/Meier/1/?distance=75

The search should be made like this:



The results will look like that:



When you download the adress file it will be a text file looking like that:

```
BEGIN:VCARD
VERSION:2.1

N;ENCODING=QUOTED-PRINTABLE;LANGUAGE=en-CZ;CHARSET=ISO-8859-2:Meier Daniel
FN;ENCODING=QUOTED-PRINTABLE;LANGUAGE=en-CZ;CHARSET=ISO-8859-2:Meier Daniel
TEL;TYPE=WORK;TYPE=VOICE:233 373 975
ADR;WORK;ENCODING=QUOTED-PRINTABLE;ENCODING=QUOTED-PRINTABLE;LANGUAGE=en-CZ;CHARSET=ISO-8859-
2:;;Kamenick=E1 652/48;Praha;Hlavn=ED m=ECsto Praha;17000;
LABEL;WORK;ENCODING=QUOTED-PRINTABLE;ENCODING=QUOTED-PRINTABLE;LANGUAGE=en-CZ;CHARSET=ISO-
8859-2:Kamenick=E1 652/48=0D=0A17000=20Praha=0D=0A

END:VCARD
```

This is probably the best way to have it clear as parseable adress-record when you download every VCARD File and parse the data into a MySQL database.

Please parse all data you see on the webpage into the MySQL database, like tel, fax, email, street, housenr, zip code, city, etc... – just EVERYTHING needs to be parsed into the database!!!

The ZLATE STRANKY Website has 2 areas:

a) Private customers (White pages directory)
b) Business Customers (Yellow pages directory)

You also should parse and get all business entries, where it's important, that you get the tel, fax, mobile nr and even the category in which business category the listing is published...

I recommend you to always download the VCF file, but also parse the data on the webpage as well... Just import it even twice like:

Name Webpage:	Meier
Name VCF File:	Meier
Firstname Webpage:	Frank
Firstname VCF File:	Frank

(so in the MySQL Database would be the same information 2 times, where it can be possible, that in the VCF File there are different information than on the webpage... what I hear...)

The business VCF Files entries look like that:

```
BEGIN:VCARD
VERSION:2.1

N;ENCODING=QUOTED-PRINTABLE;LANGUAGE=en-CZ;CHARSET=ISO-8859-2:MEDIATEL, spol. s r.o.
FN;ENCODING=QUOTED-PRINTABLE;LANGUAGE=en-CZ;CHARSET=ISO-8859-2:MEDIATEL, spol. s r.o.
TEL;TYPE=WORK;TYPE=VOICE:841 121 212
TEL;TYPE=WORK;TYPE=FAX:224 812 319
URL;TYPE=WORK:http://www.mediatel.cz
EMAIL;INTERNET:info@mediatel.cz
ADR;WORK;ENCODING=QUOTED-PRINTABLE;ENCODING=QUOTED-PRINTABLE;LANGUAGE=en-CZ;CHARSET=ISO-8859-2:;;Th=E1mova
137/16;Praha;Hlavn=ED m=ECsto Praha;18600;
LABEL;WORK;ENCODING=QUOTED-PRINTABLE;ENCODING=QUOTED-PRINTABLE;LANGUAGE=en-CZ;CHARSET=ISO-8859-2:Th=E1mova
137/16=0D=0A18600=20Praha=0D=0A

END:VCARD
```

The business entries of companies look like this:

en.zlatestranky.cz/businesses/Praha/H122589/MEDIATEL%2C+spol.+s+r.o./

Supersearch | Businesses | People | Maps

What — mediatel — Keyword, Name, Phone, Heading or Products
Where — 11000 — City, Street, Postcode
In Around — none
Find

Zlaté Stránky > M > Marketing - consultants > M > MEDIATEL, spol. s r.o.

< Back | Map | Get Directions | Save to: My Places | Send to: Outlook | Email | Mobile | Share on:

MEDIATEL, spol. s r.o.
Thámova 137/16, 18600 Praha 8 - Karlín
Tel.: 841 121 212
Fax: 224 812 319
Category: *Marketing - consultants, Hotels*

www.mediatel.cz
www.zlatestranky.cz
www.internet123.cz
www.inform.cz
www.travelguide.cz
info@mediatel.cz

Like — Be the first of your friends to like this.

Details | Ratings & Reviews

Other Contact Numbers (5)
Tel.:848 123 123
Tel.:257 102 102
Tel.:800 108 001
Tel.:724 642 624
Tel.:543 251 704
Fax: 221 807 447
View All

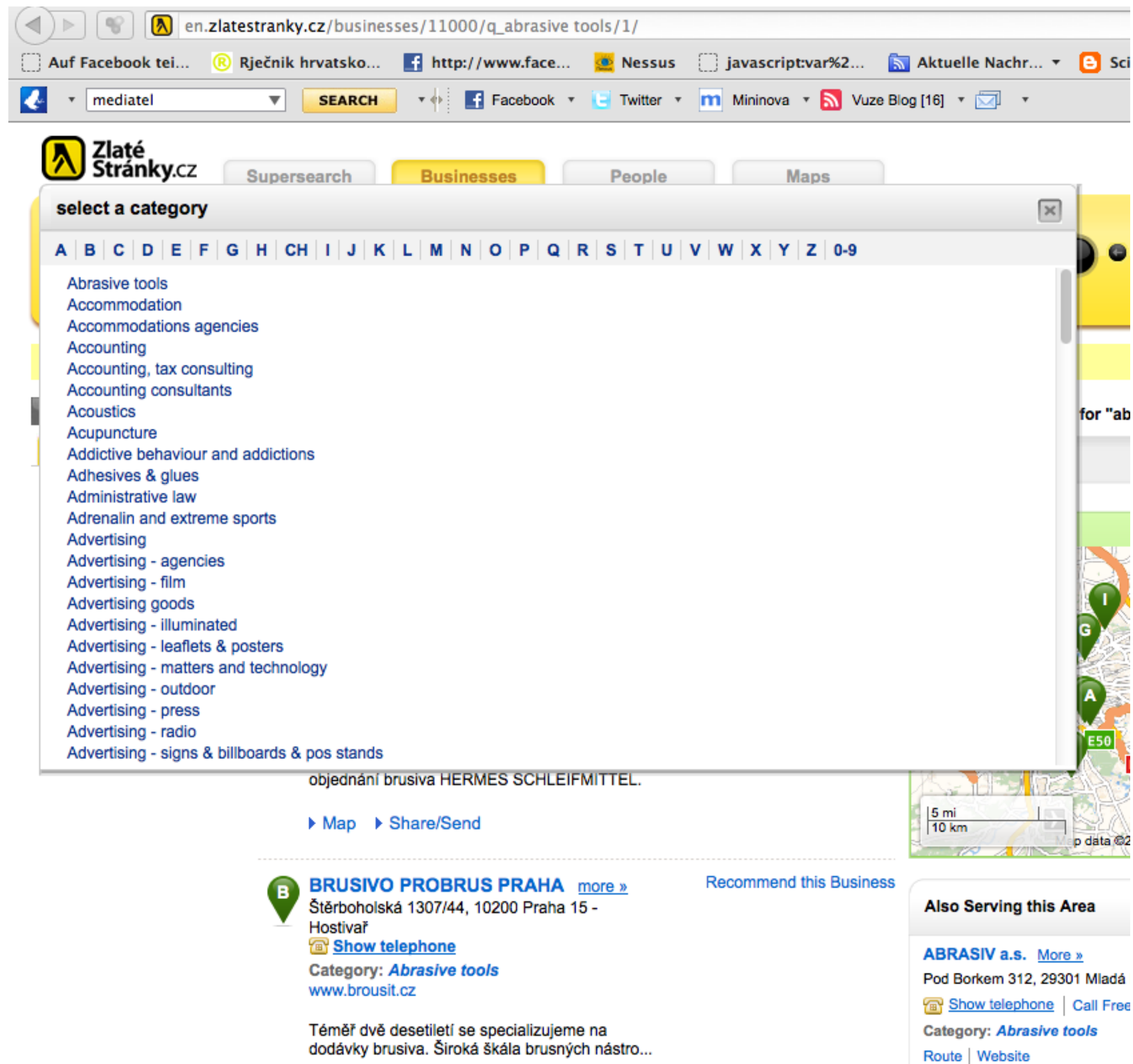
Another E-mail Addresses
jaroslav.vrtiska@travelguide.cz

Description:
Společnost Mediatel nabízí široké portfolio marketingových nástrojů pro efektivní podporu podnikání. Poskytuje unikátní mix komunikačních kanálů - online platformu www.zlatestranky.cz s řadou

See large map — Karlín — Křižíkova — Pernerova — Find Nearby

Just be sure that you import all information you can get from the webpage!!! As well as even the Czech language Description of the company as text... and make „View All“ in the script that you get all telephone numbers...

When you build up the database, please first create a own table with all business categories:

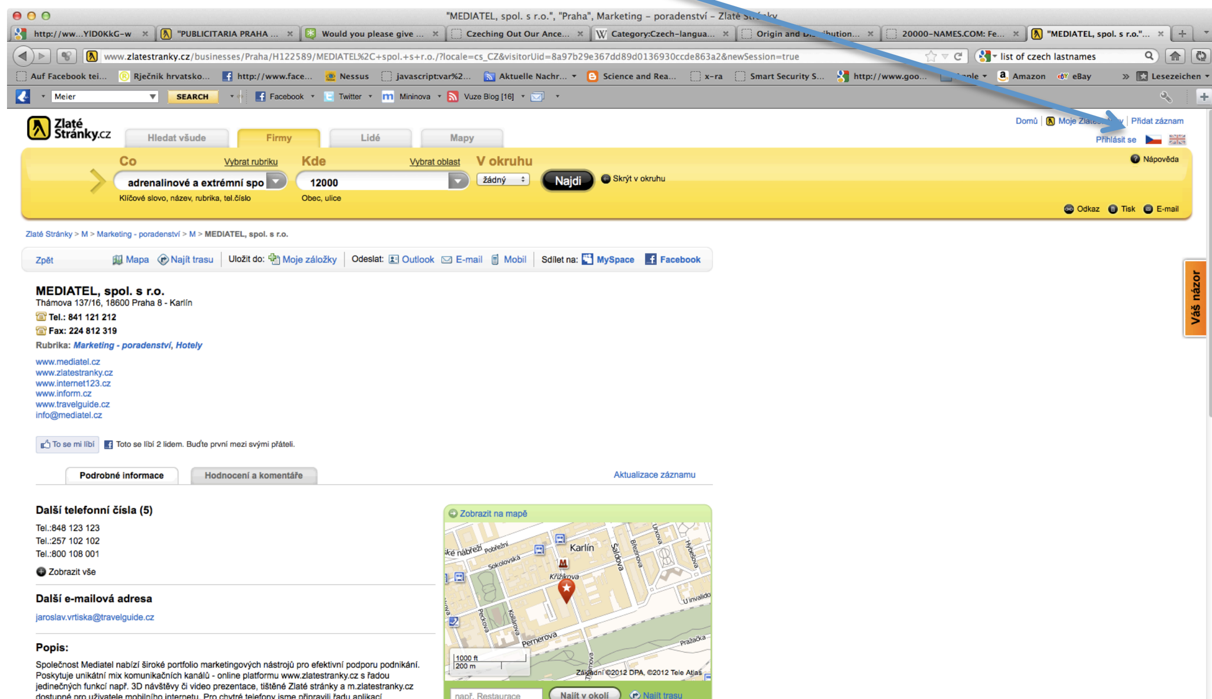


I want that you then crawl all business listings by that order:

1.) Select Business Category
2.) Select Zip code 11000, 12000, 13000, ... , 99000
And select (find all listings in 75km area...)

Multilanguage Crawling:

I want that you also crawl the Czech-language version of the entry:



TODO:

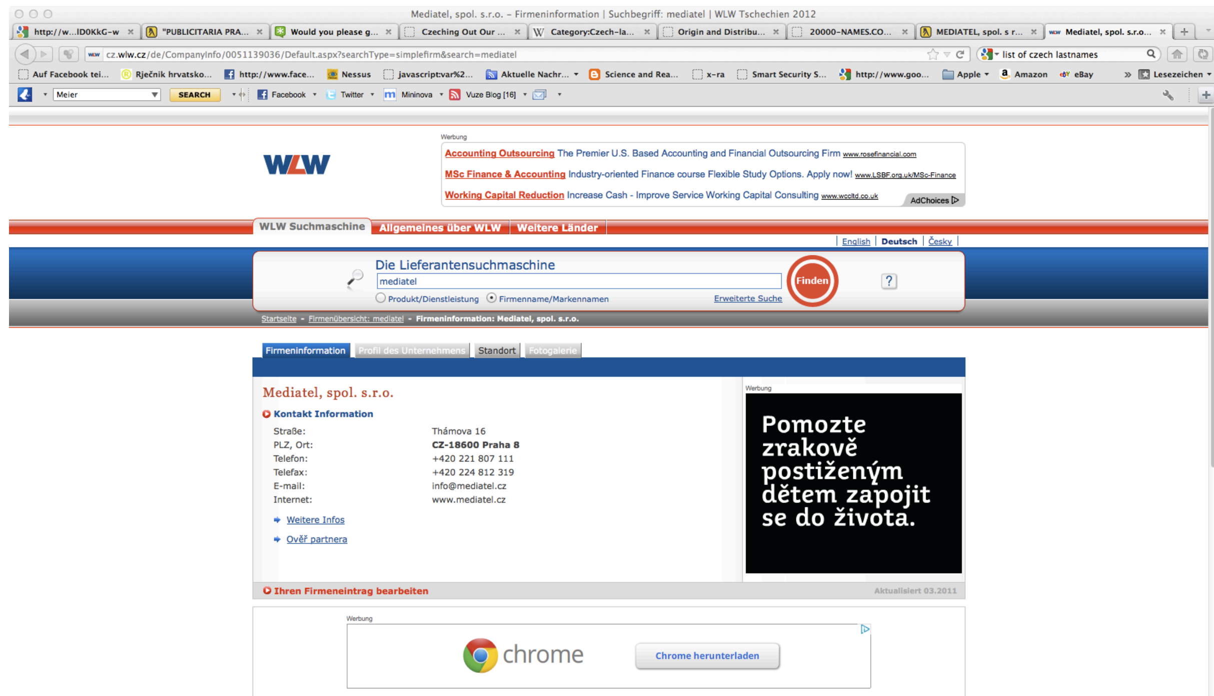
- Please add tot he relevant entry the Czech definition of the business category and als the English category
- Please crawl all data as EN (English) selection an das CZ (Czech) selection and add everything you can get from the webpage to the MySQL Database

Information:

- You don't need to download the MAP picture... this we don'T need...

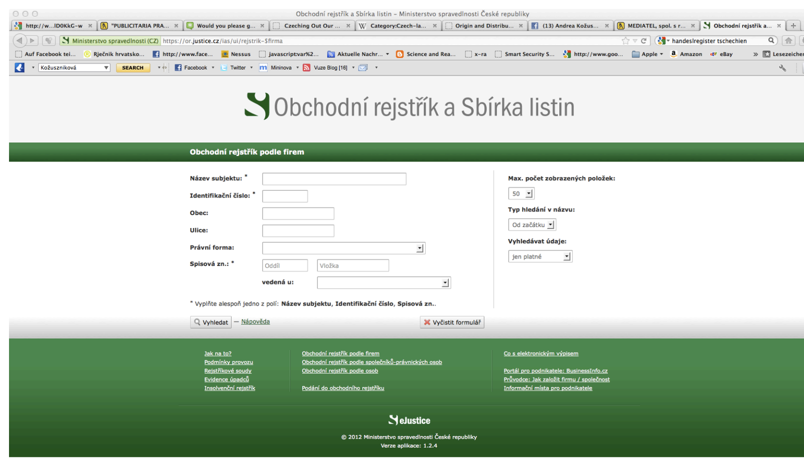
Additional databases:

Because we can also make it better, I want you after you have a list of all companies from Czech Republic and businesses, that you also check every entry on the WLW webpage:



I also want that you retreive every company in Czech Republic from the gouvernment webpage:

https://or.justice.cz/ias/ui/rejstrik-$firma



Important to take care of:

➔ Czech Republic has special characters, so you have to have them enabled in the database.... like Kožuszniková

Also parse that database:

http://w...lD0KkG-w
"PUBLICITARIA PR
www.ksv.at/KSV/1870/
Auf Facebook tei...
Rječnik hrvatsko...
Kožuszniková
SEARCH
KSV
Wir bieten
Mitgliedschaft
Servic
My KSV - Login
KSV1870
ratisBusinessSearch
Firma
PLZ
Ort
Land Tschechien
Suchen
Bonitätsauskunft bestellen
Willkomm

KSV

https://www.myksv.at/NASApp/web/gbs/ShowFreeDetail.do?searchIdbId=1&searchtype=CZ&ksvnumber=17933121

KSV1870 Login | Kunde werden | Kontakt | Sitemap | Suchen

Wir bieten | Mitgliedschaft | Services | Privatpersonen | Presse | Über uns

My KSV - Login

KSV-GratisBusinessSearch

Firma	Kožusznikovà
PLZ	
Ort	
Land	Tschechien

Suchen

Das erste Mal hier?
- Unternehmer
- Privatperson
- Your first visit at KSV

- Verwertungsbörse
- Akademie
- Jobangebote

Ihre Anfrage bei der KSV-GratisBusinessSearch für Tschechien brachte folgendes Ergebnis:

Firmenbasisinformation:

KSV-Nummer:	17933121
Name:	Andrea Kozusznikovà
Adresse:	Mezidolí 14/1344, 736 01 Havírov, Tschechische Republik
Rechtsform:	Einzelunternehmen
Branche :	Handelsvermittlung

Ausgezeichnete Bonität oder Insolvenzgefahr? Was dahinter steckt erfahren Sie nach Login.

Login Abfrage mit Kreditkarte

Zurück Neue Suche

In Kooperation mit Coface Central Europe.

Der Abfragende darf aus der KSV-GratisBusinessSearch erhaltene Daten nur für den eigenen Gebrauch verwenden. Jede darüber hinausgehende Verwertung oder Verwendung, insbesondere zu kommerziellen Zwecken, ist untersagt. Der Inhalt der KSV-GratisBusinessSearch ist urheberrechtlich geschützt (§ 40 f UrhG), Verletzungen werden zivil-und strafrechtlich verfolgt.

AGB | Impressum | Downloads |

And : - also parse that webpage database, because here you find also ICO numbers...

http://rejstrik.penize.cz/andrea-kozusznikova

Andrea Kožuszniková, IČO: 62309234, 08.04. – ARES – Peníze.cz

rejstrik.penize.cz/ares/62309234-andrea-kozusznikova

Tip redakce

OSVČ: Propad příjmů na mateřské lze odvrátit i přes úsporný balíček
Podnikatelky, které nyní nastoupily na mateřskou, čeká v příštím roce značná finanční ztráta. Mateřská jim od 1. ledna poklesne až o několik tisíc korun. I přes... Více

Čechů je málo, ať přijdou cizinci
Čechů je na trhu práce málo. Proto vláda rozhodla, že usnadní cizincům cestu za prací a firmám k náboru nových zaměstnanců. Od příštího roku by měli zahraniční pracovníci... Více

Může jít o desetitisíce! Budete po změně životního pojištění vracet odpočty z daní?
Pádným argumentem pro životní pojištění je i možnost odčítat zaplacené pojistné – ročně až 12 000 Kč – od základu daně. Ta za určitých podmínek platila, i když se... Více

Andrea Kožuszniková - IČO: 62309234 Zpět na vyhledávání

Statistický úřad ARES

TIP REDAKCE: Proč vás nezaměstnám

Datum vzniku:	31.8.1995
Obchodní firma:	Andrea Kožuszniková
Podnikatel:	Andrea Kožuszniková
Sídlo:	Havířov - Město, Mezidolí 14/1344, PSČ 736 01
Identifikační číslo:	62309234
Právní forma:	Fyzická osoba podnikající dle živnostenského zákona nezapsaná v obchodním rejstříku
Klasifikace ekonomických činností - CZ-NACE:	Zprostředkování velkoobchodu a velkoobchod v zastoupení
Institucionální sektor:	Ostatní osoby samostatně výdělečně činné
Počet zaměstnanců:	Neuvedeno

Hodnocení firmy -5 0 +5 Hodnocení: 0

Zobrazit sídlo společnosti na mapě

Informace z těchto stránek mají pouze informativní charakter. Provozovatel portálu neručí za správnost poskytnutých informací.

CELEBRITY V REJSTŘÍKU

Jsou to frajeři, nebo lupiči? Hlasujte v hitparádě o krále a největšího podvraťáka našeho obchodního rejstříku. Pokud nevidíte svého favorita, zkuste si stránku znovu načíst, objeví se další várka...

1. **Jaromír Jágr** - hokejista, mistr světa a olympijský vítěz. Hodnocení: +3
2. **Karel Gott** - zpěvák. Hodnocení: +2
3. **Petr Kellner** - podnikatel, majoritní vlastník holdingové společnosti PPF Group N.V. Hodnocení: 0
4. **Leoš Mareš** - moderátor. Hodnocení: -3
5. **Václav Kočka** - podnikatel

Whois
Please make whois lookup for every domain name in Czech Republic you find in the listings. Just make a script that for every domain name (found in email adresses or webpages) the record is retreived and stored in the database..

```
%whois -h whois.nic.cz mediatel.cz
%  (c) 2006-2012 CZ.NIC, z.s.p.o.
%
% Intended use of supplied data and information
%
% Data contained in the domain name register, as well as information
% supplied through public information services of CZ.NIC association,
% are appointed only for purposes connected with Internet network
% administration and operation, or for the purpose of legal or other
% similar proceedings, in process as regards a matter connected
% particularly with holding and using a concrete domain name.
%
% Full text available at:
% http://www.nic.cz/page/306/intended-use-of-supplied-data-and-information/
%
% See also a search service at http://www.nic.cz/whois/
%
%
% Whoisd Server Version: 3.4.0
% Timestamp: Sun Apr 08 21:28:12 2012

domain:       mediatel.cz
registrant:   SB:MEDIATEL
admin-c:      MEDIATEL-ADMIN
nsset:        NSS:MEDIATEL:1
registrar:    REG-1X
status:       paid and in zone
registered:   13.09.1996 02:00:00
changed:      03.04.2012 22:38:35
expire:       28.10.2012

contact:      SB:MEDIATEL
org:          MEDIATEL, spol. s r.o.
name:         MEDIATEL, spol. s r.o.
address:      Thamova 137/16
address:      Praha 8
address:      186 00
address:      CZ
e-mail:       webmaster@zlatestranky.cz
registrar:    REG-CT
created:      10.08.2001 22:13:00
changed:      13.09.2007 14:25:00

contact:      MEDIATEL-ADMIN
name:         Martin Smutny
address:      Thamova 137/16
address:      Praha 8
address:      186 00
address:      CZ
phone:        +420 221807403
e-mail:       msmutny@zlatestranky.cz
registrar:    REG-CT
created:      10.08.2001 22:13:00

nsset:        NSS:MEDIATEL:1
nserver:      ns.zlatestranky.cz
nserver:      au0.zlatestranky.cz
tech-c:       SB:MEDIATEL
tech-c:       MEDIATEL-ADMIN
tech-c:       MEDIATEL-TECH
registrar:    REG-CT
created:      01.10.2007 02:00:00

contact:      MEDIATEL-TECH
name:         Milos Bilwachs
address:      Thamova 137/16
address:      Praha 8
address:      186 00
address:      CZ
phone:        +420 221807240
e-mail:       mbilwachs@zlatestranky.cz
registrar:    REG-CT
created:      10.08.2001 22:13:00
```

Job:

I need somebody who can write a script which parses the webpage with the white-pages (Telephone book) data into a MySQL Database from Zlatestranky.cz and some other business-directories. I need a list of adressdata from people and companies in the Czech Republic, so I want to import from the webpage the phone-book and some other directories.

I have written a detailled instruction what I need as PDF...